PROBABILITY-WEIGHTED MOMENTS ESTIMATORS FOR TCEV PARAMETERS

Nigel Arnell and Max Beran

Institute of Hydrology

August 1988

Introduction

The Two Component Extreme Value (TCEV) distribution was advocated by Rossi et al. (1984) for use in flood frequency analysis. They presented a procedure, modified by Fiorentino et al. (1986), for estimating site and regional parameters by maximum likelihood. This note describes an alternative estimation procedure based on probability weighted moment (PWM). Details of the TCEV distribution, its fit to real data and its statistical properties, are given in Rossi et al. (1984), Arnell and Beran (1987) and Arnell and Gabriele (1988), and are not repeated here.

The TCEV distribution of annual maxima has the distribution function:

$$F(x) = \exp[- \lambda_1 e^{-x/\theta_1} - \lambda_2 e^{-x/\theta_2}] \qquad x \geq 0 \qquad .. (1)$$

or, in the standardised regional case

$$F(y) = \exp[- e^{-y} - \lambda_* e^{-y/\theta_*}] \qquad .. (2)$$

where

$$\theta_* = \theta_2/\theta_1$$

$$\lambda_* = \lambda_2/(\lambda_1^{1/\theta_*}) \qquad .. (3)$$

and

$$y = \frac{x - \theta_1' \ln\lambda_1'}{\theta_1'} \qquad .. (4)$$

θ_1' and λ_1' are site parameters, whilst θ_1, θ_2, λ_1 and λ_2 are regional parameters.

Probability-weighted Moments

Probability-weighted moments are defined as (Greenwood et al, 1979)

$$M_{p,r,s} = E\,[x^p\, F(x)^r (1 - F(x))^s] \qquad .. (5)$$

where p, r and s are real numbers. PWMs for the TCEV distribution can be derived with p = 1 and s = 0 in equation (5) (Beran et al. 1986):

$$\beta_r = PWM_r^{(1)} + \frac{\theta_1}{r+1} T_r \qquad .. (6)$$

where

$$PWM_r^{(1)} = \frac{\theta_1}{r+1} \{ \gamma + \ln \lambda_1 + \ln(r+1) \} \quad \text{. . (7)}$$

and

$$T_r = \sum_{j=1}^{\infty} (-1)^{j-1} \lambda_*^j (r+1)^{j-1/\theta_*)} \Gamma(j/\theta_*)/j! \quad \text{. . (8)}$$

γ is the Euler number.

Two different procedures for estimating TCEV parameters using PWMs have been developed, based on L-moments (Hosking, 1986) and central PWMs.

PWM Estimators Based on L-moments

U-statistics are linear combinations of PWMs (Hosking, 1986), and four define completely the TCEV:

$$L_1 = \beta_o = \theta_1 [\gamma + \ln \lambda_1 + T_o] \quad \text{. . (9a)}$$

$$L_2 = 2\beta_1 - \beta_o = \theta_1 [\ln 2 + D_1] \quad \text{. . (9b)}$$

$$L_3 = 6(\beta_2 - \beta_1) + \beta_o = \theta_1 [\ln 9/8 + 2D_2 - D_1] \quad \text{. . (9c)}$$

$$L_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_o$$
$$= \theta_1 [\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1] \quad \text{. . (9d)}$$

where T_r is as above and

$$D_r = T_r - T_{r-1}$$

These equations could be solved to yield the four TCEV parameters, but an easier solution is obtained using L-moments, which are ratios of the U-statistics (Hosking, 1986).

$$\tau = L_2/L_1 = (\ln 2 + D_1)/(\gamma + \ln\lambda_1 + T_o) \quad \text{. . (10a)}$$

$$\tau_3 = L_3/L_2 = (\ln 9/8 + 2D_2 - D_1)/(\ln 2 + D_1) \quad \text{. . (10b)}$$

$$\tau_4 = L_4/L_2 = (\ln 2^{16}/3^{10} + 5D_3 - 5D_3 - 5D_2 + D_1)/(\ln 2 + D_1) \quad \text{. . (10c)}$$

τ_3 and τ_4 are dimensionless, and are measures of skewness and kurtosis. θ_* and λ_* can be obtained using just τ_3 and τ_4 by solving the following two equations:

$$f_1 = \frac{\ln 9/8 + 2D_2 - D_1}{\ln 2 + D_1} - M3 = 0 \qquad \text{. . (11a)}$$

and

$$f_2 = \frac{\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1}{\ln 2 + D_1} - M4 = 0 \qquad \text{. . (11b)}$$

where M3 and M4 are the sample estimates of the two L-moments τ_3 and τ_4. Equations (11a) and (11b) can be solved using the Newton-Raphson method:

$$u^{(k+1)} = u^{(k)} + du \quad ; \quad t^{(k+1)} = t^{(k)} + dt$$

$$u = \theta_* \ln \lambda_*$$

$$t = 1/\theta_* \qquad \text{. . (12)}$$

where

$$du = - \left(\frac{df_2}{dt} f_1 - \frac{df_1}{dt} f_2 \right) / \left(\frac{df_1}{du} \frac{df_2}{dt} - \frac{df_1}{d} \frac{df_2}{du} \right) \qquad \text{. . (13a)}$$

$$dt = - \left(- \frac{df_2}{du} f_1 + \frac{df_1}{du} f_2 \right) / \left(\frac{df_1}{du} \frac{df_2}{dt} - \frac{df_1}{dt} \frac{dt_2}{du} \right) \qquad \text{. . (13b)}$$

The derivatives are

$$\left.\begin{aligned}
\frac{df_1}{du} &= \frac{1}{g} [2D_{2\,u} - (1+h_1)D_{1,u}] \\
\frac{df_1}{dt} &= \frac{1}{g} [2D_{2,t} - (1+h_1)D_{1,t}] \\
\frac{df_2}{du} &= \frac{1}{g} [5D_{3,u} - 5D_{2,u} + (1-h_2)D_{1,u}] \\
\frac{df_2}{dt} &= \frac{1}{g} [5D_{3,t} - 5D_{2,t} + (1-h_2)D_{1,t}]
\end{aligned}\right\} \qquad \text{. . (14)}$$

where

$$
\left.
\begin{aligned}
D_{r,u} &= \frac{dT_r}{du} - \frac{dT_{r-1}}{du} \; ; \qquad D_{r,t} = \frac{dT_r}{dt} - \frac{dT_{r-1}}{dt} \\
\frac{dT_r}{du} &= t \sum_{j=1}^{\infty} \frac{(-1)^{j-1} \, e^{jut} \, \Gamma(jt) \, (r+1)^{j(1-t)}}{(j-1)!} \\
\frac{dT_r}{dt} &= [u - \ln(r+1)] \sum_{j=1}^{\infty} \frac{(-1)^{j-1} \, e^{jut} \, \Gamma(jt) \, (r+1)^{j(1-t)}}{(j-1)!} \\
&\quad + \sum_{j=1}^{\infty} e^{jut} \, \frac{\Gamma(jt) \, \psi(jt)(r+1)^{j(1-t)}}{(j-1)!} \\
g &= \ln 2 + D_1 \; ; \\
h_1 &= \frac{\ln 9/8 + 2D_2 - D_1}{g} \; ; \\
h_2 &= \frac{\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1}{g}
\end{aligned}
\right\} \quad \ldots (15)
$$

A regional TCEV distribution can be fitted using the regional average of the sample estimates of M3 and M4 (weighted by record length), giving regional θ_* and λ_*, and hence a regional frequency curve in terms of the normalised variate

$$
Y = \frac{x - \theta_1' \ln\lambda_1'}{\theta_1'} \quad \ldots (16)
$$

θ_1' and λ_1' are site parameters. This regional curve can be converted into a curve showing the T-year flood as a multiple of the mean $\bar{x}$ using a regional average λ_1'. Site estimates of λ_1' are obtained by solving equation (10a) with site estimates M2 of τ, and these site values can be averaged to give $\bar{\lambda}_1$. Alternatively a regional average M2 can be used to determine $\bar{\lambda}_1$.

Constrained TCEV distributions can be easily fitted to individual samples, where θ_*, λ_* and $\bar{\lambda}_1$ are fixed. θ_1' can be obtained explicitly from Equation (9a), and θ_2' and λ_2' are determined determined from θ_*, λ_* θ_1' and $\bar{\lambda}_1$.

PWM Estimators Using Central PWM's

This version of the PWM estimation procedure was developed by Fabio Rossi, and uses central PWMs defined as

$$\beta_r' = E[\,(x - \beta_o)F^r] \qquad \text{.. (17)}$$

Central PWMs are linear functions of the more usual non-central PWMs β_r:

$$\beta_r' = \beta_r - \frac{\beta_o}{r+1}$$

The first two central PWMs are

$$\beta_o' = 0 \qquad \text{.. (18a)}$$

$$\beta_1' = \beta_1 - \frac{\beta_o}{2} = \theta_1[\ln 2 + T_1 - T_o]/2 \qquad \text{.. (18b)}$$

$$\beta_2' = \beta_2 - \frac{\beta_o}{3} = \theta_1[\ln 3 + T_2 - T_o]/3 \qquad \text{.. (18c)}$$

$$\beta_s' = \beta_s - \frac{\beta_o}{4} = \theta_1[\ln 4 + T_3 - T_o]/4 \qquad \text{.. (18d)}$$

(Note that $L_2 = 2\beta_1'$

$$L_3 = 6(\beta_2' - \beta_1') \qquad \text{.. (19)}$$

and

$$L_4 = 20\beta_3' - 30\,\beta_2' + 12\beta_1' \qquad)$$

TCEV parameters θ_* and λ_* are estimated using ratios of these central PWMs:

$$\frac{3}{2}\,(\beta_2'/\beta_1') = \frac{\ln 3 + T_2 - T_o}{\ln 2 + T_1 - T_o} \qquad \text{.. (20a)}$$

$$2(\beta_3'/\beta_1') = \frac{\ln 4 + T_3 - T_o}{\ln 2 + T_1 - T_o} \quad \text{. . (20b)}$$

and, as before, the solution can be found using a Newton-Raphson method. In the regional case θ_* and λ_* are estimated using regional averages of sample values B2/B1 and B3/B1. Equations 10b and 10c differ from 20a and 20b only in the numerator, and the central PWM ratios are therefore related to the L-moments by

$$\frac{3}{2} \; (\beta_2'/\beta_1') = \tau_3 \; \frac{\ln 3 + T_2 - T_o}{\ln 9/8 + 2D_2 - D_1} \quad \text{. . (21a)}$$

$$2(\beta_3'/\beta_1') = \tau_4 \; \frac{\ln 4 + T_3 - T_o}{\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1} \quad \text{. . (21b)}$$

Regional average values of $\bar{\lambda}_1'$ can be obtained by averaging the site estimates computed from:

$$2(\beta_1'/\beta_o) = \frac{\ln 2 + T_1 - T_o}{\ln\lambda_1 + T_o + \gamma} \quad \text{. . (22)}$$

where the site estimate B1'/B0 is used to estimate β_1'/β_o. Equation (22) is equivalent to Equation 10(a). Site estimates of θ_1 can be obtained from Equation (18b) using site estimates of the central PWMs.

Practical Application

Both versions of the PWM estimator for the regional TCEV distribution are much simpler to apply than the maximum likelihood procedure currently used.

Firstly, there is no need to start with at-site estimates of θ_1' and λ_1', since the dimensionless ratios M3 and M4 (and B3/B1 and B2/B1) describe both the distribution of x and its linear transform Y. This is a major advantage over the maximum likelihood procedure, as it has been found that estimation of θ_1' and λ_1' is often the most difficult stage. The estimation of the parameters θ_* and λ_* is also much quicker, as there is no need to compute the values of the likelihood function and its derivatives over a large sample, and it is of course not necessary to repeat the process with new estimates of at-site θ_1' and λ_1'.

The studies so far have used unbiased estimates of PWMs:

$$\hat{\beta}_r = n^{-1} \sum_{j=1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_j \quad \ldots (23)$$

Regional TCEV distributions were fitted to data from 58 UK stations using both the L-moments and central PWM estimators, and compared with a distribution fitted by maximum likelihood. Parameter estimates are shown in Table 1, and Table 2 shows regional quantiles in terms of both reduced variate Y and the ratio of the T-year flood to the mean $\bar{x}$.

Table 1 Regional TCEV parameters, UK data

	TCEV-PWM L-moments	TCEV-PWM central	TCEV-MLE
θ_*	6.0253	6.0533	4.4542
λ_*	0.0117	0.0116	0.0293
$\bar{\lambda}_1$	16.0002	15.9977	18.4146
outlier probability	0.011	0.011	0.026

Table 2 Regional frequency curves, TCEV distribution

Return period	TCEV-PWM L-moments	TCEV-PWM central	TCEV-MLE
(a) Y_T quantiles			
20	3.117	3.116	3.291
50	4.242	4.239	4.624
100	5.268	5.265	6.011
500	10.722	10.725	11.965
1000	14.837	14.859	15.043
(b) $x_T/\bar{x}$			
20	1.725	1.724	1.719
50	2.054	1.053	2.088
100	2.354	2.354	2.473
500	3.951	3.953	4.122
1000	5.156	5.163	4.975

It is clear from the tables that the two PWM estimators are very similar, and this is to be expected given that they are linearly related: in fact, they should be identical. The PWM estimators give very different parameter estimates to those from the maximum likelihood fit, although the difference between the quantiles is not so great.

During the fitting it was found that the choice of starting estimates of θ_* and λ_* for the Newton-Raphson procedure was important, and that with some starting points the algorithm did not converge. However, when it did converge it always reached the same solution. The problem was eliminated by incorporating a random starting point, with new starting points repeatedly chosen until the algorithm converged.

Some simulation experiments have been conducted with the L-moment version of the PWM estimator, using synthetic regions with 40 stations each of 40 years drawn from TCEV distributions based on fits to UK and Italian data (to enable comparisons with results of the TCEV-MLE experiments presented in Arnell and Gabriele (1988)). The most important finding was that for a high proportion of synthetic regions the TCEV-PWM algorithm failed to reach a solution - the proportions failing were 40.8% for the UK-based parent and 7.4% for the Italian-based parent (the TCEV-MLE procedure reached a solution for all of the synthetic regions). These failures were not (on the whole) due to inadequate starting points, but reflect the limited range of combinations of τ_3 and τ_4 feasible with the TCEV distribution. Figure 1 shows θ_* and λ_* plotted against L-moments τ_3 and τ_4, and it is clear that there is a large number of combinations of τ_3 and τ_4 which cannot be accommodated by the TCEV distribution. It was found that the vast majority of synthetic regions which did not yield a solution had sample estimates of M3 and M4 which fell outside the 'feasible space' of the TCEV distribution. This has also been found with some real-world flood and rainfall data, and suggests a significant weakness in the TCEV-PWM procedure.

D. Jones (pers. comm.) has developed approximations for the bounds implied in Fig. 1. The upper bound is given by:

$$\tau_4' \simeq \frac{\tau_3 \left(\ln \left(2^{15}/3^{10}\right)\right) + \ln 3^{12}/2^{19}}{\ln(3^2/2^4)} \qquad \text{. . (24)}$$

and the lower bound is defined by:

$$\tau_3' \simeq \frac{2\ [\ \ln 3/2 + E_1(3\lambda_*) - E_1(2\lambda_*)\ ]}{\ln 2 + E_1(2\lambda_*) - E_1(\lambda_*)} - 1 \qquad \text{. . (25)}$$

$$\tau_4' \simeq \frac{5\ [\ \ln 8/9 + E_1(4\lambda_*) - 2E(3\lambda_*) + E_1(2\lambda_*)\ ]}{\ln 2 + E_1(2\lambda_*) - E_1(\lambda_*)} + 1$$

where $E_1(\ )$ is the exponential integral. The TCEV is only defined for τ_3 and τ_4 above the line defined by (τ_3', τ_4').

Table 3 shows the mean and standard deviation of estimates of dimensionless $x_T/\bar{x}$ quantiles for both parents and for both ML and PWM estimates: it must be remembered that a solution was not found for all PWM repetitions.

The simulation results show that the TCEV-PWM method tends to produce lower growth curves (except for higher quantiles with the Italian parent and low quantiles with the UK parent) with a variability similar to that of the TCEV-MLE procedure (although variability is less with TCEV-PWM at high quantiles with the UK parent). Overall, the TCEV-PWM algorithm does not seem to show a dramatic improvement over the TCEV-MLE procedure, and any reduced bias tends to be offset by higher standard deviation.

Table 3 *Mean and standard deviation of growth factors* $x_T/\bar{x}$*, 40 stations with 40 years each, 500 repetitions.*

	return period				
	20	50	100	500	1000
(a) TCEV-1 (UK-based parent)					
true	1.645	1.981	2.334	3.793	4.532
	mean growth factors				
TCEV-MLE	1.636	1.968	2.310	3.654	4.532
TCEV-PWM	1.658	1.979	2.277	3.388	4.062
	standard deviation				
TCEV-MLE	0.020	0.054	0.122	0.536	0.752
TCEV-PWM	0.024	0.056	0.106	0.423	0.691
(b) TCEV-2 (Italian-based parent)					
true	2.037	2.721	3.324	4.794	5.431
	mean growth factors				
TCEV-MLE	2.032	2.689	3.260	4.649	5.251
TCEV-PWM	2.017	2.644	3.221	4.717	5.375
	standard deviation				
TCEV-MLE	0.044	0.095	0.170	0.407	0.515
TCEV-PWM	0.048	0.098	0.160	0.443	0.596

Summary

From an operational point of view, the TCEV-PWM algorithm is much quicker and easier to apply than the TCEV-MLE algorithm, but restrictions in the range of feasible combinations of PWM ratios mean that the algorithm does not always converge: this has been found with both real and synthetic data. The PWM and ML estimators can give quite different parameter estimates when applied to the same data set, although estimated quantiles are less different.

Acknowledgements

The work presented in this note was funded by the Ministry of Agriculture, Fisheries and Food.

References

Arnell, N.W. and Gabriele, S., 1988. The performance of the two component extreme value distribution in regional flood frequency analysis. *Water Resources Research 24*, 879-887.

Arnell, N.W. and Beran. M.A., 1987. Testing the suitability of the TCEV distribution for regional flood estimation. In: Singh, V.P. (ed) *Regional Flood Frequency Analysis*, Redel, Dordrecht, 159-175.

Beran, M.A., Hosking, J.R.M. and Arnell, N.W., 1986. Comment on 'Two-component extreme value distribution for flood frequency analysis'. *Water Resources Research 22*, 263-266.

Fiorentino, M., Versace, P. and Rossi, F., 1985. Regional flood frequency analysis using the two-component extreme value distribution. *Hydrol. Sci. J. 30*, 51-64.

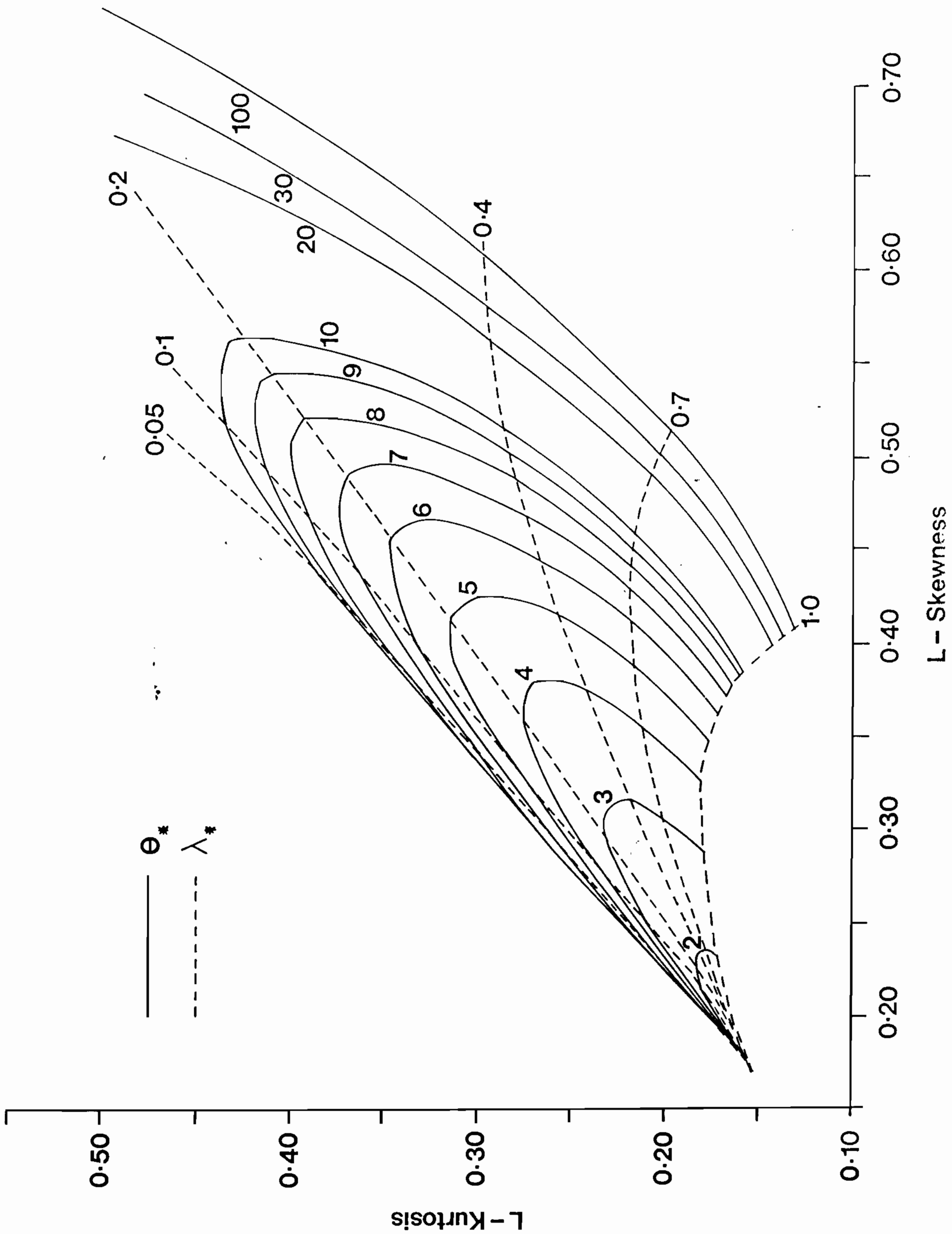
Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research 15*, 1049-1054.

Hosking, J.R.M., 1986. *The theory of probability weighted moments.* IBM Research Report 12210.

Rossi, F., Fiorentino, M. and Versace, P., 1984. Two-component extreme value distribution for flood frequency analysis. *Water Resources Research* *20*, 847-856.

NA/vw
19.8.88

Figure 1. *Variation of* θ_* *and* λ_* *with L-skewness and L-kurtosis.*



The demand for long-term scientific capabilities concerning the resources of the land and its freshwaters is rising sharply as the power of man to change his environment is growing, and with it the scale of his impact. Comprehensive research facilities (laboratories, field studies, computer modelling, instrumentation, remote sensing) are needed to provide solutions to the challenging problems of the modern world in its concern for appropriate and sympathetic management of the fragile systems of the land's surface.

The **Terrestrial and Freshwater Sciences** Directorate of the Natural Environment Research Council brings together an exceptionally wide range of appropriate disciplines (chemistry, biology, engineering, physics, geology, geography, mathematics and computer sciences) comprising one of the world's largest bodies of established environmental expertise. A staff of 550, largely graduate and professional, from four Institutes at eleven laboratories and field stations and two University units provide the specialised knowledge and experience to meet national and international needs in three major areas:

★

Land Use and Natural Resources

★

Environmental Quality and Pollution

★

Ecology and Conservation